Speech Processing 2016/17

1st Test

March 23rd 2016

Please identify this form with your name and student number in the reserved spaces in the two sheets. Wrong answers to True/False questions will be penalized. The phonetic symbols should use the SAMPA alphabet, Lisbon accent.

Name:	
Number:	
Group number:	
Place of birth:	

1. Classify as True (T) or False (F) on the **left** side of each item.
   (a) The main lobe of a rectangular window has a smaller bandwidth than Hamming or Hanning windows of the same length.
   (b) In unvoiced fricatives, it is easy to distinguish formants.
   (c) The auditory canal has a fairly broad resonance around 1 kHz.
   (d) The auditory neurons have a latency period of 1-3ms in which, having once fired, cannot fire again.
   (e) The covariance method of linear prediction involves the computation of reflection coefficients.

2. What can a person do to modify the shape of his / her own glottal pulses?

   _____

3. What characteristic can we measure in waveforms / spectrograms to help us distinguish unvoiced from voiced plosives, and how it differs in these two classes?

   _____

4. Fill in the blanks:
   In _____ spectrograms, horizontal striations are visible in _____ regions. A typical window length for this type of spectrograms is _____ ms.

5. Consider a pole-zero filter. What are the conditions that allow an easy computation of its complex cepstrum from its real cepstrum?

   _____

6. Write the expression of the cepstrum of a voiced segment as a sum of 4 components, indicating the meaning of each component.

7. Consider a sine wave of maximum/minimum amplitude +/-1, and frequency F0=500 Hz, sampled at Fs=8000 Hz.

   (a) Compute the Zero Crossing Rate as the number of zero crossings in a 10 ms interval.

   (b) Will this number be the same, if the sinewave includes a DC component?

   (c) ZCR may be used to distinguish between two classes of segments. Indicate which.

8. A sustained vowels file, similar to the one used in Lab 1, has been recorded with 8kHz, 16 bits per sample, duration 15.0 s. Pitch was estimated with the autocorrelation method, using 30 ms windows, updated every 10 ms. The file has been edited to leave just 0.6 s between consecutive vowels and in the beginning and end of the utterance. (Please indicate only the computations).

   (a) How many values were computed and stored in file vowels.myf0 ?

   (b) How many of these values were zero (approximately)?

   (c) What was the average duration of a vowel?

9. Consider the following sentence in European Portuguese:
   A esperança média de vida em Portugal aumentou. As mulheres tendem a viver mais anos que os homens.

   (a) Write the broad phonetic transcription.

   (b) Which of the following classes of sounds is less represented (as a percentage of total number of sounds of that class)?
   Oral vowels, nasal vowels, nasal consonants, oral glides, nasal glides, fricatives, plosives, vibrants, laterals.

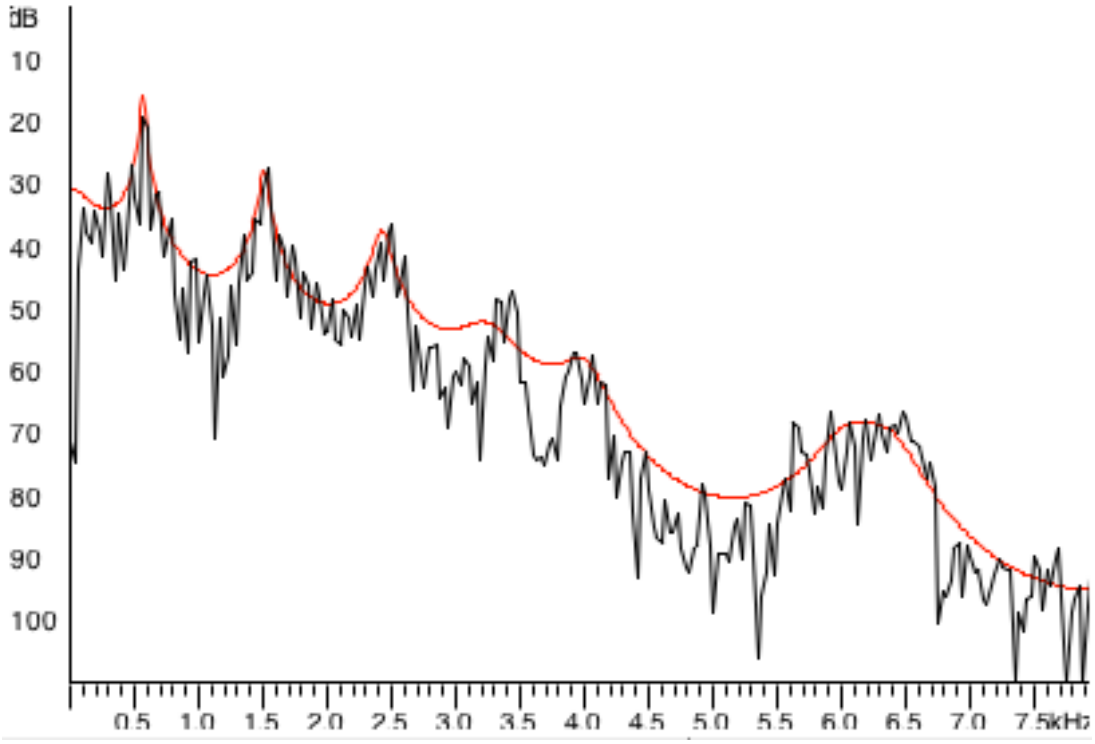


Figure 1

10. Figure 1 presents the short-time Fourier transform (magnitude) of a vowel segment, including its LPC spectral envelope.

    (a) Indicate approximate values for F0, F1, F2 and F3 (Hz).
    F0= _____ F1= _____
    F2 = _____ F3 = _____

    (b) To which of the 4 vowels of the extremes of the vowel triangle, or the neutral vowel does it correspond?
    _____

    (c) Does it correspond to a male/female/child voice?
    _____

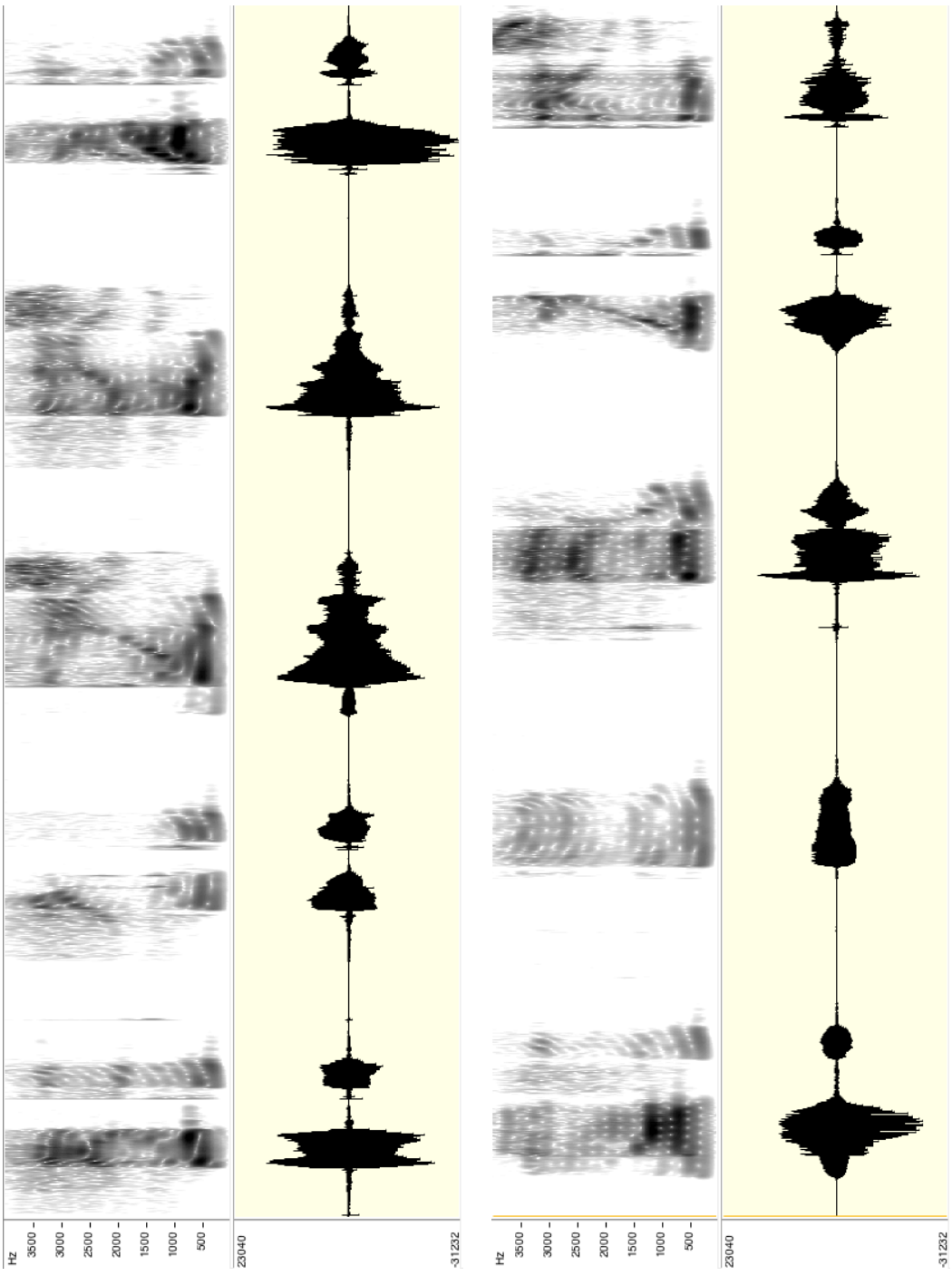
    (d) Indicate a plausible range of values for the order of the linear prediction analysis used in the computation of the envelope.
    _____

    (e) Sketch superimposed in the figure the corresponding magnitude of short-time Fourier transform of the linear prediction residual.

    (f) Which of the four first formants has poles furthest away from the circumference (unit radius) ?
    _____

    (g) The above segment corresponds to a non-pathological type of phonation characterized by irregular f0. What are the possible phonation types?
    _____

11. Identify the 10-digit sequence by inspecting the spectrograms and waveforms in the next page. There are no repeated digits. The recordings correspond to telephone speech.



20 = 1.5 + 1.0 + 1.0 + 1.5 + 1.0 + 1.0 + 2.0 + 2.0 + 2.5 + 3.5 + 3.0
Number: _____